



Document Summary



New
Search



Help

[Preview Claims](#)

[Preview Full Text](#)

[Preview Full Image](#)

Email Link: 

Document ID: J P 05-101102 A2

Title: RETRIEVAL DEVICE

Assignee: MATSUSHITA ELECTRIC IND CO LTD

Inventor: MURAI KATSUMI

US Class:

Int'l Class: G06F 15/40 A

Issue Date: 04/23/1993

Filing Date: 10/03/1991

Abstract:

PURPOSE: To reduce the capacity of a file used for a presearch for retrieving all sentences and to improve the efficiency of the retrieval by retrieving only part of a list which contains characters of a retrieved character string and checking whether or not the characters of the retrieved character string are all present or not at the end of main body retrieval.

CONSTITUTION: A character string retrieval circuit 5 retrieves document data read out of a secondary storage device 1 to a data memory circuit 6 and a circuit 7 which controls the secondary storage device 1, data memory circuit 6, and character string retrieval circuit 5 by receiving a request from a host computer 8, is provided to generate the table of the character string having characters of previously recorded document data as an entry. This table is used to check whether or not there is the same character string as the retrieved character string and when there is the same character string, a corresponding identification name (corresponding to the entry of a recording place of the secondary storage) is obtained. Consequently, the data amount of a communication between a host computer 8 and the secondary storage device 1 is decreased to speed up the retrieval.

(C)1993,JPO&Japio

Legal Notices

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平5-101102

(43)公開日 平成5年(1993)4月23日

(51)Int.Cl.⁵

G 0 6 F 15/40

識別記号

5 0 0 D 7060-5L

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数17(全 21 頁)

(21)出願番号 特願平3-256215

(22)出願日 平成3年(1991)10月3日

(71)出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72)発明者 村井 克己

大阪府門真市大字門真1006番地 松下電器

産業株式会社内

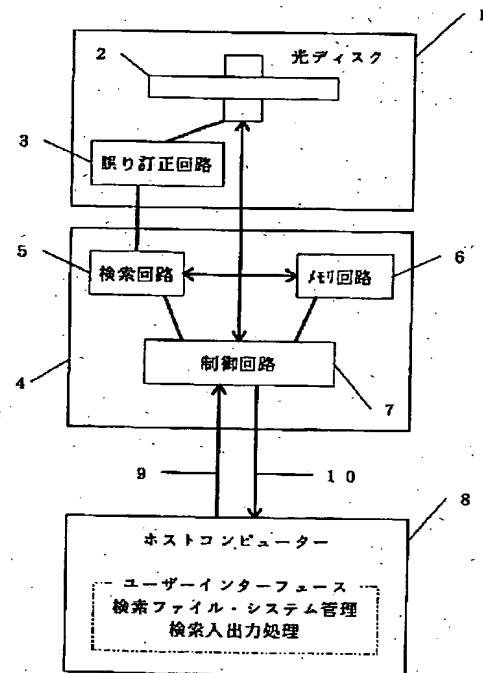
(74)代理人 弁理士 小銀治 明 (外2名)

(54)【発明の名称】 検索装置

(57)【要約】

【目的】 大量の文書データを蓄えた2次記憶装置から検索用のインデックス情報を付与することなしに要求された文書データを引き出してくる全文検索方式に基づく検索装置に関するもので、特に全文検索のプリサーチに用いるファイルの容量を小容量化し検索の効率をよくする。

【構成】 要約ファイルとして、全文ファイルの文字並びから記録場所を連想する表を作っておき、この表を用いて絞り込んだ本文についてのみ検索を行う。検索語入力時にやはり文字並びについて調べ、連想表と対応する文字の一部をキーとして表の一部の検索を行う。



【特許請求の範囲】

【請求項1】(a) 2次記憶装置と、(b)データメモリ回路と、(c)データパターン検出回路と、(d)検索要求を外部より受け付けて前記2次記憶装置、前記データメモリ回路、前記データパターン検出回路とを制御する制御回路とを具備し、(e)前記2次記憶装置の記録媒体に検索対象となる文書データを区別して蓄える手段と、(f)前記の文書データの区別された領域毎に、前記2次記録装置の記録媒体上の記録場所情報と、前記記録場所情報に対応した識別名称とを与える手段と、(g)前記識別名称から前記記録場所情報を得る手段と、(h)前記の文書データを調べて文字の長さ M ($M \geq 1$)の任意に設定した特定文字列が存在するかどうか検索しても一致箇所が見つかった場合には、一致発生箇所の前か後ろの文字の長さ N ($N \geq 1$)の文字列と前記の識別名称とを得て、前記の文字の長さ M の任意に設定した特定文字列の種類を索引情報として、前記の識別名称と、前記の一致発生箇所の前か後ろの文字の長さ N の文字列からなる文字列情報との1対組の情報を得て表として予め順次記録しておく手段と、(i)要求者から出された j 個 ($j \geq 1$)の検索文字列と要求者から出された各々の検索文字列間の検索文字列論理式を検索条件として設定する手段と、(j) j 個の前記検索文字列のうちの一つの検索文字列中に存在するところの文字の長さ M の検索文字列部分文字列、および文字の長さ N の検索文字列部分文字列前後文字列の組を、 k 組 ($k \geq 1$)だけ得て、前記検索文字列部分文字列を索引として前記検索文字列部分文字列前後文字列を前記の表の中から探し、対応する検索文字列 i の部分要素に対応する識別名称を順次 k 個だけ得る手段と、(k)前記の手段「(j)」により k 組の数だけ繰り返して得た検索文字列 i ($i = 1, \dots, j$)の部分要素に対応する識別名称群の中から、 k 組のグループにすべて存在 (AND) するところの検索文字列 i に対応する識別名称を検索候補として選び出す手段と、(l)検索要求のあった時点で前記の検索候補である検索文字列 i に対応する識別名称から前記「(g)」の手段により記録場所情報を得て、前記2次記憶装置中の前記文書データ中から絞り込んだ検索対象本文を前記データメモリに読み取る手段とからなり、前記の表を参照して絞り込んだ検索対象本文を前記データメモリに読み、前記のデータパターン検出回路を適用して、前記の j 個の検索文字列が存在してかつ要求装置から出された前記の検索文字列論理式を満たすかどうかをチェックした後、条件を満たした検索対象本文あるいは検索対象本文に関する情報を要求者に提示することを特徴とする検索装置。

【請求項2】(a) 2次記憶装置と、(b)データメモリ回路と、(c)データパターン検出回路と、(d)検索要求を外部より受け付けて前記2次記憶装置、前記データメモリ回路、前記データパターン検出回路とを制御する制御回路とを具備し、(e)前記2次記憶装置の記録媒体に検索対

象となる文書データをファイル単位あるいは任意の量単位に区別して蓄える手段と、(f)前記の文書データの区別された単位毎に、前記2次記録装置の記録媒体上の記録場所情報と、前記記録場所情報に対応した識別名称とを与える手段と、(g)前記識別名称を参照して前記記録場所情報を得る手段と、(h)任意の1文字を予めグループ分けして文字グループと、グループ内で符号化するグループ内文字符号を得る手段と、(i)前記の文書データを調べて文字の長さ M ($M \geq 1$)の任意に設定した特定文字列とこの前か後ろの任意に設定した1文字の特定の前記文字グループが存在するかどうか検索し、もし一致箇所が見つかった場合には、前記特定の文字グループのグループ内文字記号と、前記の1文字の前か後ろの文字の長さ N ($N \geq 0$)の文字列と一致箇所の前記の識別名称とを得て、前記の文字の長さ M の任意に設定した特定文字列と、その前か後ろの1文字の任意に設定した特定の前記文字グループの種類を索引情報として、前記の識別名称と、前記特定の文字グループのグループ内文字符号および前記の1文字の前か後ろの文字の長さ N の文字列とからなる文字列情報とを得て、一致箇所の情報の表として予め順次記録しておく手段と、(j)要求者から出された j 個 ($j \geq 1$)の検索文字列と要求者から出された各々の検索文字列間の検索文字列論理式を検索条件として設定する手段と、(k) j 個の前記検索文字列のうちの1検索文字列中に存在するところの文字の長さ M の検索文字列部分文字列、およびその前か後の1文字の特定の検索文字列文字グループと文字グループ内文字符号、および1文字の前記特定の検索文字列文字の前か後ろの文字の長さ N の検索文字列部分文字列前後文字列の組を、 k 組 ($k \geq 1$)だけ得て、前記検索文字列部分文字列および前記検索文字列文字グループを索引として前記検索文字列文字グループ内符号と文字の長さ N の前記検索文字列部分文字列前後文字列を前記の表の中から探し、対応する検索文字列 i の部分要素に対応する識別名称を順次 k 個だけ得る手段と、(l)前記の手段「(k)」により k 組の数だけ繰り返して得た検索文字列 i の部分要素に対応する識別名称群の中から、 k 組のグループにすべて存在 (AND) するところの検索文字列 i に対応する識別名称を検索候補として選び出す手段と、(m)検索要求のあった時点で前記の検索候補である検索文字列 i に対応する識別名称から前記「(g)」の手段により記録場所情報を得て、前記2次記憶装置中の前記文書データ中から絞り込んだ検索対象本文を前記データメモリに読み取る手段とからなり、前記の表を参照して絞り込んだ検索対象本文を前記データメモリに読み、前記のデータパターン検出回路を適用して、前記の j 個の検索文字列が存在してかつ要求装置から出された前記の検索文字列論理式を満たすかどうかをチェックした後、条件を満たした検索対象本文あるいは検索対象本文に関する情報を要求者に提示することを特徴とする検索装

置。

【請求項3】(a)2次記憶装置と、(b)データメモリ回路と、(c)データパターン検出回路と、(d)検索要求を外部より受け付けて前記2次記憶装置、前記データメモリ回路、前記データパターン検出回路とを制御する制御回路とを具備し、(e)前記2次記憶装置の記録媒体に検索対象となる文書データを区別して蓄える手段と、(f)前記の文書データの区別された領域毎に、前記2次記憶装置の記録媒体上の記録場所情報と、前記記録場所情報に対応した識別名称とを与える手段と、(g)前記識別名称から前記記録場所情報を得る手段と、(h)1文字を予めグループ分けして文字グループを得る手段と、(i)前記の文書データを調べて文字の長さ M ($M \geq 1$)の任意に設定した特定文字列が存在するかどうか検索しても一致箇所が見つかった場合には、一致発生箇所の前か後ろの文字の長さ N ($N \geq 1$)の文字列と、前記の長さ N の文字列の前か後ろの1文字の属する前記文字グループと、前記の識別名称とを得て、前記の文字の長さ M の任意に設定した特定文字列の種類を索引情報として、前記の識別名称と、前記の一致発生箇所の前か後ろの文字の長さ N の文字列および前記の文字グループとからなる文字列情報とを得て、1対組の情報の表として予め順次記録しておく手段と、(j)要求者から出された j 個 ($j \geq 1$)の検索文字列と要求者から出された各々の検索文字列間の検索文字列論理式を検索条件として設定する手段と、(k)前記 j 個の検索文字列のうちの一つの検索文字列中に存在するところの前記の文字の長さ M の検索文字列部分文字列、および前記の文字の長さ N の検索文字列部分文字列前後文字列、および前記の長さ N の文字列の前か後ろの1文字の属する前記文字グループとの組を、 k_i 組 ($k_i \geq 1$)だけ得て、長さ M の前記検索文字列部分文字列を索引として、長さ N の前記検索文字列部分文字列前後文字列および前記の長さ N の文字列の前か後ろの1文字の属する前記文字グループの、1対組の情報を前記の表の中から探し、対応する検索文字列 i の部分要素に対応する識別名称を順次 k_i 個だけ得る手段と、(k)前記の手段「(j)」により k_i 組の数だけ繰り返して得た検索文字列 i の部分要素に対応する識別名称群の中から、 k_i 組のグループにすべて存在 (AND) するところの検索文字列 i に対応する識別名称を検索候補として選り出す手段と、(l)検索要求のあった時点で前記の検索候補である検索文字列 i に対応する識別名称から「(g)」の手段により前記記録場所情報を得て、前記2次記憶装置中の前記文書データ中から絞り込んだ検索対象本文を前記データメモリに読み取る手段とからなり、前記の表を参照して絞り込んだ検索対象本文を前記データメモリに読み、前記のデータパターン検出回路を適用して、前記の j 個の検索文字列が存在してかつ要求装置から出された前記の検索文字列論理式を満たすかどうかをチェックした後、条件を満たした検索対象本文あるいは検索対

象本文に関係する情報を要求者に提示することを特徴とする検索装置。

【請求項4】1対組の情報の表には、索引情報として、文字の長さ M の任意に設定した特定文字列と、前記特定文字列の前か後ろの1文字の任意に設定した特定の第1の文字グループの種類を用い、内容としては前記の1文字の特定の第1の文字グループのグループ内文字符号、および前記の1文字の特定の第1の文字グループの前か後ろの文字の長さ N の文字列と、前記長さ N の文字列の前か後ろの前記とは別の1文字の属する第2の文字グループとからなる文字列情報とを得てこれらを文字情報として、該当する文書データの識別名称と共に1対組の情報として順次書き込み、また j 個の前記検索文字列のうちの1検索文字列中に存在するところの文字の長さ M の検索文字列部分文字列、およびその前か後の1文字の特定の検索文字列の第1の文字グループと文字グループ内文字符号、および1文字の前記特定の検索文字列文字の前か後ろの文字の長さ N の検索文字列部分文字列前後文字列と、前記長さ N の文字列の前か後ろの前記とは別の1文字の属する第2の文字グループとからなる文字列情報とを、 k_i 組 ($k_i \geq 1$)だけ得て、前記検索文字列部分文字列および前記検索文字列文字グループを索引として、前記検索文字列の第1の文字グループ内符号と、文字の長さ N ($N \geq 0$)の前記検索文字列部分文字列前後文字列と前記の第2の文字グループとを前記の表の中から探し、対応する検索文字列 i の部分要素に対応する識別名称を順次 k_i 個だけ得る手段を有する請求項2または請求項3記載の検索装置。

【請求項5】(a)2次記憶装置と、(b)データメモリ回路と、(c)データパターン検出回路と、(d)検索要求を外部より受け付けて前記2次記憶装置、前記データメモリ回路、前記データパターン検出回路とを制御する制御回路とを具備し、(e)前記2次記憶装置の記録媒体に検索対象となる文書データを区別して蓄える手段と、(f)前記の文書データの区別された領域毎に、前記2次記憶装置の記録媒体上の記録場所情報と、前記記録場所情報に対応した識別名称とを与える手段と、(g)前記識別名称から前記記録場所情報を得る手段と、(h)前記の文書データのある時点で調べて文字あるいは文字の並びを索引として各文字並びが前記2次記憶装置のどこに書かれているかの対応関係を表として予め順次記録しておく手段と、(i)要求者から出された j 個 ($j \geq 1$)の検索文字列と要求者から出された各々の検索文字列間の検索文字列論理式を検索条件として設定する手段と、(j) j 個の前記検索文字列のうちの一つの検索文字列中に存在するところの文字あるいは文字の並びを索引として前記の表の中から探し、検索文字列 i ($i = 1 \dots j$)の文字の並びに対応する識別名称を順次 k_i 個だけ得る手段と、(k)前記の k_i 個の検索文字列 i の文字の並びに対応する識別名称群の中から、 k_i 組のグループにすべて存在

(AND) するところの検索文字列 i に対応する識別名称を検索候補として選び出す手段と、(l) 検索要求のあった時点で前記の検索候補である検索文字列 i に対応する識別名称から前記「(g)」の手段により記録場所情報を得て、前記 2 次記憶装置中の前記文書データ中から絞り込んだ検索対象本文を前記データメモリに読み取る手段と、(m) 検索要求のあった時点で、以前に前記表を作成した時点以降に新たに追加された文書データについて、未だ前記の表が得られていない場合は、前記の新たに追加された前記文書データをもすべて検索対象として前記「(l)」に引き続き前記データメモリに読み取る手段と、(n) 要求により前記の新たに追加された文書データについて、未だ前記の表が得られていない前記の新たに追加された文書データを調べて文字の並びが前記 2 次記憶装置のどこに書かれているかの対応関係を表として順次追加記録していく手段からなり、検索要求のあった時には前記の表を参照して絞り込んだ検索対象本文を前記データメモリに読むとともに、前記の未だ前記の表が得られていない前記の新たに追加された文書データをも、前記データメモリに読み込み、前記のデータパターン検出回路を適用して、前記の j 個の検索文字列が存在してかつ要求装置から出された前記の検索文字列論理式を満たすかどうかをチェックした後、条件を満たした検索対象本文あるいは検索対象本文に関する情報を要求者に提示することを特徴とする検索装置。

【請求項 6】(a) 2 次記憶装置と、(b) データメモリ回路と、(c) データパターン検出回路と、(d) 検索要求を外部より受け付けて前記 2 次記憶装置、前記データメモリ回路、前記データパターン検出回路とを制御する制御回路とを具備し、(e) 前記 2 次記憶装置の記録媒体に検索対象となる文書データを区別して蓄える手段と、(f) 前記の文書データの区別された領域毎に、前記 2 次記録装置の記録媒体上の記録場所情報と、前記記録場所情報に対応した識別名称とを与える手段と、(g) 前記識別名称から前記記録場所情報を得る手段と、(h) 前記の文書データを要求のあったある時点で調べて文字あるいは文字の並びを索引として各文字並びが前記 2 次記憶装置のどこに書かれているかの対応関係を表として予め順次記録しておく、かつ前記表の記録形式は前記表を形成する各々の文字あるいは文字並びの索引毎に前記 2 次記憶装置の連続したセクタに記録し、かつ前記各々の文字あるいは文字並びの索引毎に引かれる表データの間には予測される増加分数、あるいは一定数の無記録セクタを挿入しておくか、あるいはまた前記各々の文字あるいは文字並びの索引毎に引かれる表データを複数集めた領域毎の間には予測される増加分数、あるいは一定数の無記録セクタを挿入しておく手段と、(i) 要求者から出された j 個 ($j \geq 1$) の検索文字列と要求者から出された各々の検索文字列間の検索文字列論理式を検索条件として設定する手段と、(j) j 個の前記検索文字列のうちの一つの検

索文字列中に存在するところの文字あるいは文字の並びを索引として前記の表の中から捜し、検索文字列 i ($i = 1, \dots, j$) の文字の並びに対応する識別名称を順次 k_i 個だけ得る手段と、(k) 前記の k_i 個の検索文字列 i の文字の並びに対応する識別名称群の中から、 k_i 組のグループにすべて存在 (AND) するところの検索文字列 i に対応する識別名称を検索候補として選び出す手段と、(l) 検索要求のあった時点で前記の検索候補である検索文字列 i に対応する識別名称から前記「(g)」の手段により記録場所情報を得て、前記 2 次記憶装置中の前記文書データ中から絞り込んだ検索対象本文を前記データメモリに読み取る手段と、(m) 文書データが新たに追加された時、要求により前記追加文書データを調べて文字あるいは文字の並びを索引として各文字並びが前記 2 次記憶装置のどこに書かれているかの対応関係を表データから得て、過去に作成した文字並びの表の格納されている前記 2 次記憶装置の最終セクタを読みだして追加記録して表を形成し、もし前記セクタに書き込む余地がなくなった場合は前記の無記録セクタに書き込むか、さらにそれでも書き込む余地がなくなった場合には新たに最終セクタとして別の場所に無記録セクタを確保して記録しておくとともにその旨を識別情報として記録しておく手段とからなり、前記の表を参照して絞り込んだ検索対象本文を前記データメモリに読み、前記のデータパターン検出回路を適用して、前記の j 個の検索文字列が存在してかつ要求装置から出された前記の検索文字列論理式を満たすかどうかをチェックした後、条件を満たした検索対象本文あるいは検索対象本文に関する情報を要求者に提示することを特徴とする検索装置。

【請求項 7】セクタに追加して書き込む余地がなくなった場合は新たに別のセクタに記録しておくとともにその旨を識別情報として記録しておいて得た前記の最終セクタを、要求に従って表データの文字あるいは文字並びの索引毎に連続するように、また前記各々の文字あるいは文字並びの索引毎に引かれる表データの間には予測される増加分数、あるいは一定数の無記録セクタを挿入しておくか、あるいはまた前記各々の文字あるいは文字並びの索引毎に引かれる表データを複数集めた領域毎の間には予測される増加分数、あるいは一定数の無記録セクタを挿入しておくように再配置する処理手順を設けて実行することを特徴とする請求項 6 記載の検索装置。

【請求項 8】(a) 2 次記憶装置と、(b) データメモリ回路と、(c) データパターン検出回路と、(d) 検索要求を外部より受け付けて前記 2 次記憶装置、前記データメモリ回路、前記データパターン検出回路とを制御する制御回路とを具備し、(e) 前記 2 次記憶装置の記録媒体に検索対象となる文書データを区別して蓄える手段と、(f) 前記の文書データの区別された領域毎に、前記 2 次記録装置の記録媒体上の記録場所情報と、前記記録場所情報に対応した識別名称とを与える手段と、(g) 前記識別名称か

ら前記記録場所情報を得る手段と、(h)前記の文書データを要求のあったある時点で前記データパターン検出回路を適用して記録内容を調べて文字あるいは文字の並びが2次記憶装置のどこに書かれているかの対応関係を表として予め順次記録しておく手段と、(i)要求者から出されたj個($j \geq 1$)の検索文字列と要求者から出された各々の検索文字列間の検索文字列論理式を検索条件として設定する手段と、(j)j個の前記検索文字列のうちの一つの検索文字列中に存在するところの文字あるいは文字の並びを索引として前記の表の中から探し、検索文字列i($i = 1, \dots, j$)の文字の並びに対応する識別名称を順次ki個だけ得る手段と、(k)前記のki個の検索文字列iの文字の並びに対応する識別名称群の中から、ki組のグループにすべて存在(AND)するところの検索文字列iに対応する識別名称を検索候補として選び出す手段と、(l)検索要求のあった時点で前記の検索候補である検索文字列iに対応する識別名称から前記「(g)」の手段により記録場所情報を得て、前記2次記憶装置中の前記文書データ中から絞り込んだ検索対象本文を前記データメモリに読み取る手段とからなり、検索要求のあった時には前記の表を参照して絞り込んだ検索対象本文を前記データメモリに読み込み、前記のデータパターン検出回路を適用して、前記のj個の検索文字列が存在してかつ要求装置から出された前記の検索文字列論理式を満たすかどうかをチェックした後、条件を満たした検索対象本文あるいは検索対象本文に関する情報を要求者に提示することを特徴とする検索装置。

【請求項9】 データパターン検出回路は前記データメモリのデータに対して読みだしアドレスを与えて指定された固定ビット数毎に一致検出処理を行い、かつ前記アドレス提示はデータに対して連続的ではなく指定されたビット数離れたデータに対してのみ一致検出処理を行うことを特徴とする請求項8記載の検索装置。

【請求項10】 (a)2次記憶装置と、(b)データメモリ回路と、(c)データパターン検出回路と、(d)検索要求を外部より受け付けて前記2次記憶装置、前記データメモリ回路、前記データパターン検出回路とを制御する制御回路とを具備し、(e)前記2次記憶装置の記録媒体に検索対象となる文書データを区別して蓄える手段と、(f)前記の文書データの区別された領域毎に、前記2次記憶装置の記録媒体上の記録場所情報と、前記記録場所情報に対応した識別名称とを与える手段と、(g)前記識別名称から前記記録場所情報を得る手段と、(h)前記の文書データを調べて文字あるいは文字の並びを索引とし、表の内容としては周辺の各文字あるいは文字並びを第1の固定ビットあるいはバイト数幅で表して、また各識別名称を第2の固定ビットあるいはバイト数幅で表して得た組を用意し、かつ記録場所情報位置は昇順あるいは降順で表して、前記記録場所情報を数値で表した場合の最大値の桁数より小さい下位の桁数で表記すると共に、前記下

位桁が桁上がりする場合には、前記の第1の固定ビットあるいはバイト数と同じ長さの定義されていない未定義記号と、第2の固定ビットあるいはバイト数と同じ長さの前記記録場所情報の上位桁か疑似データを与えるところの前記の表を作成して、予め順次記録しておく手段と、(i)要求者から出されたj個($j \geq 1$)の検索文字列と要求者から出された各々の検索文字列間の検索文字列論理式を検索条件として設定する手段と、(j)j個の前記検索文字列のうちの一つの検索文字列中に存在するところの文字あるいは文字の並びを索引として前記の表の中から探し、検索文字列i($i = 1, \dots, j$)の文字の並びに対応する識別名称を順次ki個だけ得る手段と、(k)前記のki個の検索文字列iの文字の並びに対応する識別名称群の中から、ki組のグループにすべて存在

(AND)するところの検索文字列iに対応する識別名称を検索候補として選び出す手段と、(l)検索要求のあった時点で前記の検索候補である検索文字列iに対応する識別名称から前記「(g)」の手段により記録場所情報を得て、前記2次記憶装置中の前記文書データ中から絞り込んだ検索対象本文を前記データメモリに読み取る手段とからなり、前記の表を参照し、前記表の索引から引用された内容を昇順あるいは降順に走査して絞り込んだ検索対象本文を前記データメモリに読み、前記のデータパターン検出回路を適用して、前記のj個の検索文字列が存在してかつ要求装置から出された前記の検索文字列論理式を満たすかどうかをチェックした後、条件を満たした検索対象本文あるいは検索対象本文に関する情報を要求者に提示することを特徴とする検索装置。

【請求項11】 (a)2次記憶装置と、(b)データメモリ回路と、(c)データパターン検出回路と、(d)検索要求を外部より受け付けて前記2次記憶装置、前記データメモリ回路、前記データパターン検出回路とを制御する制御回路とを具備し、(e)前記2次記憶装置の記録媒体に検索対象となる文書データを区別して蓄える手段と、(f)前記の文書データの区別された領域毎に、前記2次記憶装置の記録媒体上の記録場所情報と、前記記録場所情報に対応した識別名称とを与える手段と、(g)前記識別名称から前記記録場所情報を得る手段と、(h)要求を外部より受け付けて前記の文書データの文字をある特定の予め定めたビットあるいはバイト単位に正規化して前記データメモリ回路に供給する手段と、(i)前記の文書データを前記2次記憶から複数回読み出して前記データメモリに送り、前記読み出しのうちの各一回では前記のデータパターン検出回路が前記データメモリを走査して、予め設定した文字群が含まれる文字の部分パターンをテンプレートとして与え、前記文字部分パターン検出の後に前記データメモリの前記検出箇所周辺のデータを読みだし詳細に調べて前記の該当文字群の個々の文字の特定を行い、さらに文字並びを含む詳細情報を前記の特性された文字群の個々に対応させたメモリ領域に仕分けして格納

する手段と、(j)前記2次前記のデータパターン検出回路は前記データメモリの文書データに対してアドレスを前記の予め定めたビットあるいはバイト単位に提示して読みだし、かつ前記の予め定めたビットあるいはバイト数よりは少ない文字部分パターン毎にデータを一致比較するように前記データパターン比較回路を設定し、前記のデータメモリの走査する際に前記データパターン検出回路が設定した前記文字部分パターンのテンプレートとの一致を検出したら、前記の「(i)」の手段に基づいて仕分けされた詳細情報から前記の文字部分パターンのテンプレートを文字の一部とする文字群あるいは文字並び群の個々を特定して索引とし、前記の特定された個々の文字あるいは文字並び近傍の文字情報と前記識別情報を表の内容として予め順次記録しておく手段と、(k)要求者から出されたj個(j \geq 1)の検索文字列と要求者から出された各々の検索文字列間の検索文字列論理式を検索条件として設定する手段と、(l)j個の前記検索文字列のうちの一つの検索文字列中に存在するところの文字あるいは文字の並びを索引として前記の表の中から探し、検索文字列i(i=1...j)の文字の並びに対応する識別名称を順次ki個だけ得る手段と、(m)前記のki個の検索文字列iの文字の並びに対応する識別名称群の中から、ki組のグループにすべて存在(AND)するところの検索文字列iに対応する識別名称を検索候補として選び出す手段と、(n)検索要求のあった時点で前記の検索候補である検索文字列iに対応する識別名称から前記「(g)」の手段により記録場所情報を得て、前記2次記憶装置中の前記文書データ中から絞り込んだ検索対象本文を前記データメモリに読み取る手段とからなり、検索要求のあった時に、前記の表を参照して絞り込んだ検索対象本文を前記データメモリに読み込み、前記のデータパターン検出回路を適用して、前記のj個の検索文字列が存在してかつ要求装置から出された前記の検索文字列論理式を満たすかどうかをチェックした後、条件を満たした検索対象本文あるいは検索対象本文に関する情報を要求者に提示することを特徴とする検索装置。

【請求項12】(a)2次記憶装置と、(b)データメモリ回路と、(c)データパターン検出回路と、(d)検索要求を外部より受け付けて前記2次記憶装置、前記データメモリ回路、前記データパターン検出回路とを制御する制御回路とを具備し、(e)前記2次記憶装置の記録媒体に検索対象となる文書データを区別して蓄える手段と、(f)前記の文書データの区別された領域毎に、前記2次記録装置の記録媒体上の記録場所情報と、前記記録場所情報に対応した識別名称とを与える手段と、(g)前記識別名称から前記記録場所情報を得る手段と、(h)前記の文書データを調べて文字の並びが前記2次記憶装置のどこに書かれているかの対応関係を表として予め順次記録しておく手段と、(i)要求者から出されたj個(j \geq 1)の検

索文字列と要求者から出された各々の検索文字列間の検索文字列論理式を検索条件として設定する手段と、(j)j個の前記検索文字列のうちの一つの検索文字列中に存在するところの文字あるいは文字の並びを索引として前記の表の中から探し、検索文字列i(i=1...j)の文字の並びに対応する識別名称を順次ki個だけ得る手段と、(k)前記のki個の検索文字列iの文字の並びに対応する識別名称群の中から、ki組のグループにすべて存在(AND)するところの検索文字列iに対応する識別名称を検索候補として選び出す手段と、(l)要求者から出された1つの検索文字列を要求ホスト装置を經由して受け取った時点で、前記の索引表の中から索引捜した結果が出るまで前記要求ホスト装置を待たせる状態にせず、即座に要求を受け付けて終了し、前記要求ホスト装置が他の作業をしている間に前記の表の中から対応する前記識別名称を得て、結果を前記要求ホストに送らずに蓄えておき、もし前記要求ホストから結果が既に出ているかの問い合わせがあれば回答し、結果が出ていれば前記要求ホストに結果を転送する手段と、(m)要求者が前記j個の検索文字列の第i番目を前記要求ホストに入力している間にi-1番目の既に受け取った検索文字列について前記「(l)」の手段によって前記の識別名称を求める手段と、(n)検索文字列を要求者から前記要求ホストを經由して受け取り、前記要求ホストで前記各々の検索文字列間の検索文字列論理式が満たされているかをチェックして前記識別名称および前記記録場所情報を得るか、あるいはまた前記要求ホストから前記各々の検索文字列間の検索文字列論理式を受け取って検索条件を満たした前記識別名称、あるいはさらに前記記録場所情報を得て、最終の検索文字列を前記要求ホストから受け取った時点で、前記2次記憶装置中の前記文書データ中から絞り込んだ検索対象本文を前記データメモリに読み取る手段と、(o)前記の表を参照して絞り込んだ検索対象本文を前記データメモリに読み、前記のデータパターン検出回路を適用して、前記のj個の検索文字列が存在してかつ要求装置から出された前記の検索文字列論理式を満たすかどうかをチェックした後、条件を満たした検索対象本文あるいは検索対象本文に関する情報を要求者に提示することを特徴とする検索装置。

【請求項13】2次記憶装置内に、データメモリ回路と、前記データメモリ回路をアクセス可能なデータパターン検出回路とを内蔵し、検索要求のあった時点で前記のデータパターン検出回路を適用して、前記のデータメモリに読み込まれた前記のj個の検索文字列が存在しているかどうかをチェックして要求装置に条件を満たした検索対象本文あるいは検索対象本文に関する情報を要求装置に送るか、あるいはまた前記のj個の検索文字列が存在してかつ要求装置から出された前記の検索文字列論理式を満たすかどうかをチェックした後、条件を満たした検索対象本文あるいは検索対象本文に関する情報を

要求装置に送ったのち要求者に提示することを特徴とする請求項1、2、3、5、6、8、10、11、または12記載の検索装置。

【請求項14】 文字あるいは文字の並びを索引として周辺の各文字並びが2次記憶装置のどこに書かれているかの対応関係を表として予め順次記録しておく際に、2次記憶内の文書データの区別された領域内に重複して存在していないかを調べる手段を設け、重複分については表データとして記載しないことを特徴とする請求項1、2、3、5、6、8、10、11、12、または13記載の検索装置。

【請求項15】 文字あるいは文字の並びを索引として周辺の各文字並びが2次記憶装置のどこに書かれているかの対応関係を表として予め順次記録しておく際に、2次記憶内の文書データの区別された領域内に重複して存在していないかを調べる手段を設け、重複分については表データとして記載しないと共に、各文字重複数を前記の2次記憶装置の文書データの区別された領域を指定されただけ集計し、前記の表の索引であるところの文字あるいは文字の並びから引かれる周辺の各文字並びに対応した頻度数を記録する手段を設けた請求項14記載の検索装置。

【請求項16】 独立の名称を持つ文献が予め定められた範囲の文字数以下ならば複数件まとめて前記とは別の識別名称を与えて、また独立の名称を持つ文献が予め定められた範囲の文字数以上ならば複数件に分割して前記とは別の識別名称を与えて、前記の識別名称を有する文書データを2次記憶装置の区別された領域として、識別名称と対応させて前記2次記憶装置の記録媒体上の記録場所情報を得る手段と、前記各々の識別名称から前記のものの文献の独立の名称を知りまた前記文献の前記2次記憶装置内の記録場所情報を得る手段とを設けて、検索要求のあった時には表を参照して絞り込んだ検索対象本文をデータメモリに読むとともに、前記のものの文献の前記2次記憶装置内の記録場所情報を得て、前記のデータパターン検出回路を適用して、前記のj個の検索文字列が存在するかつ要求装置から出された前記の検索文字列論理式を満たすかどうかを前記のものの文献単位にチェックした後、条件を満たした検索対象本文あるいは検索対象本文に関係する情報を要求者に提示することを特徴とする請求項1、2、3、5、6、8、10、11、12、13、14、または15記載の検索装置。

【請求項17】 k_i 個の検索文字列 i ($i=1, \dots, j$)の文字の並びに対応する識別名称群の中から、 k_i 組のグループに少なくとも予め定めた数あるいは割合だけすべて存在するところの検索文字列 i に対応する識別名称を検索候補として選び出す手段を備え、検索要求のあった時に、前記の表を参照して絞り込んだ検索対象本文を前記データメモリに読み込み、前記のデータパターン検出回路を適用して、前記のj個の検索文字列に曖昧

さを許して存在するかつ要求装置から出された前記の検索文字列論理式を満たすかどうかをチェックした後、条件を満たした検索対象本文あるいは検索対象本文に関係する情報を要求者に提示することを特徴とする請求項1、2、3、5、6、8、10、11、12、13、14、15、または16記載の検索装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、大量の文書データを蓄えた2次記憶装置から検索用のインデックス情報を付与することなしに要求された文書データを引き出してくる全文検索方式を基本とした検索装置に関するものである。

【0002】

【従来の技術】近年、ワードプロセッサやパーソナルコンピュータの普及により大量の文書データが仕事場や家庭に於いて流通利用される状況になってきた。この大量の文書データを整理して有効に利用していくために、大容量データベースと高速検索マシンが研究開発されてきた。しかし、従来の検索マシンでは検索用にインデックス情報を付ける必要があり、データ量が増大するにつれてこのインデックス付け作業に大変な労力が必要となってきた。これに対して、前記インデックス情報をつける作業を必要としない方法として、前記インデックス情報なしにこの大量の文書データの中からの的確かつ高速に所望の文書データを探し出すことができる全文検索方式に基づく検索装置が発表されている。例えば、1970年スロトニック (Slotnick, D.L.) が提案したロジック・パー・トラックディスクは、2次記憶装置の一種であるディスクの各ヘッドに検索専用のプロセッサを付加し、検索条件を満足した情報だけをホストコンピュータに転送することにより検索の高速化を試みたものであり、具体的な装置としてトロント大学のRAP等が実現している。一方、全文検索用テキストサーチマシン (電子情報通信学会技術研究報告・データ工学89-38) は、2次記憶装置の複数化、文字成分表と凝縮本文という二種類の要約ファイルを用いた階層型プリサーチ方式を用いることによって検索の高速化を試みている。

【0003】

【発明が解決しようとする課題】しかしながら前記ロジック・パー・トラックディスクにおいては、全文検索を行おうとすると2次記憶装置の一種であるディスク全体を検索する必要があり、前記2次記憶装置に対するデータの入出力に必要な時間が多くなり、検索の高速化が難しくなるという問題があった。また、前記全文検索用テキストサーチマシンでは、この2次記憶装置に対するデータの入出力に必要な時間を少なくすることを目的として2次記憶装置の複数化と全文を凝縮した要約ファイルによるプリサーチ方式を用いているが、どうしても前記要約ファイルが大きくなってしまいうという欠点があつ

た。

【0004】本発明ではこれら従来の装置においては大規模なハードウェアを必要とし、また検索速度的にも十分でなかったのに対して、小さなハードウェアと効率的なプリサーチ方式を採用することによって安価で高速な検索ができる検索装置を提供することを目的とする。

【0005】

【課題を解決するための手段】上記問題点を解決するために本発明の検索装置は、2次記憶装置に蓄えられた文書データを読み出し、適当な文書量毎に識別名称を付与し、かつその文書データの内容を調べ、すべての文字（あるいは文字並び：Mは1文字以上）を先頭とし、引き続く文字（あるいは文字並び：Nは1文字以上）がどのようなものであるかを分析する。そしてこの時ある文書量毎に長さMの文字を索引としMの長さの文字に引き続く長さNの文字とこれらの文字並びが含まれている文書の識別名称を表の内容として予め記述しておく。次に検索要求者から出されたj個の検索文字列（例えば検索単語）とそれらの検索論理式を受け付け、それら検索文字列を調べてすべての文字（あるいは文字並び：Mは1文字以上）を先頭とし、引き続く文字（あるいは文字並び：Nは1文字以上）がどのようなものであるかを分析し、文字Mを索引として先ほどの表を引き、表の内容が一致するものの識別名称をki個入手する。そしてj個の検索文字列の各文字並びがすべて含まれているもの（AND）の識別名称を持つ文書を検索候補として選び、2次記憶中の実際の記録場所を得て、バッファメモリ（データメモリ）に読み込む。そして文字列検索ハードウェアに先ほどのj個の検索文字列をセットして、実際に検索文字列の文字並びが検索候補とした文書に存在するかどうかを確かめる。さらに与えられた論理式関係を満足するかどうかをも確認した後、条件を満たした検索対象本文あるいは検索対象本文に関係する情報を要求者に提示することになる。

【0006】本発明は、基本的に先行技術である全文検索用テキストサーチマシンと同様にプリサーチ方式を採用したものであるが（上記では表と記述）、本発明においてはすべて検索した場合は時間のかかってしまう全プリサーチファイルを検索の対象とせず、検索文字列の文字並びの含まれる表の一部分のみを検索する。本文検索を終了した時点ではじめて検索文字列のすべての文字並びが存在するかどうかチェックされる構成となっている。

【0007】

【作用】本発明によれば上記のように、2次記憶装置からデータメモリ回路に読み出された文書データが、文字列検索回路によって検索され、ホストコンピュータからの要求を受け付けて前記2次記憶装置、データメモリ回路、文字検索回路を制御する回路を装備して、予め記録文書データの文字をエントリとする文字並びの表を作

っておき、この表を引いて検索文字列と同一の文字並びがあるかどうか調べ、もしあったなら対応する識別名称（2次記憶の記録場所のエントリに対応）を得る。このようにすればホストコンピュータと2次記憶装置間の通信データ量が削減されて検索の高速化が可能となる。すなわち全文を凝縮したプリサーチファイルとしての文字並びの表を用いることで、要約ファイルの記録の容量を小さくでき効率の良い検索を行うことができる。従って効果的な全文検索を行うことができる。

【0008】

【実施例】以下本発明の実施例を図面を用いて詳細に説明する。

【0009】図1は本発明の検索装置の構成図である。図1において、1は大量の文書データを蓄えておく2次記憶装置であって文書データの本文ファイルと検索時に使用する連想表ファイルを記録する光ディスクドライブであり、内部には2の光ディスク媒体と、3の誤り訂正回路を装備している。また4は1の2次記憶装置と8のホストコンピュータの間におかれる処理回路である。5は検索動作を行うデータパターン検出回路であり、6は光ディスクドライブから読み出したデータを一時的に蓄えておくデータメモリ回路であり、7はホストコンピュータからの検索要求に対して前記記録媒体やデータメモリ回路また文字列検索回路を制御するところの、マイクロコントローラ及びインターフェイスを含む制御回路であり、8はユーザーからの検索要求の受け入れてユーザーへの検索結果の出力を行い前記2次記憶装置に検索要求を送り前記2次記憶装置から検索結果を受け取る働きをするホストコンピュータであり、9は前記ホストコンピュータから前記2次記憶装置に送られる検索要求データの流れであり、10は前記2次記憶装置から前記ホストコンピュータに送られる、連想表検索結果の流れと最終的に表示される検索結果の流れである。

【0010】以上の様に構成された検索装置において、図2及び図15を用いて検索動作について説明する。予め文書データ本文は1の2次記憶装置に既に蓄えられており、FAT（ファイルアロケーションテーブル）、ディレクトリ等のファイル管理システム下にあるとする。

【0011】予め9のホストコンピュータにより文書データの内容は図2の手順に従って分析される。まずステップ1でユーザーから8のホストコンピュータに文書データの分析要求が出される。すなわちホストコンピュータは識別名称として2次記憶媒体のファイル単位の物理的配置の順に端から番号をふられた文書データファイルをすべて読み出してホストコンピュータのメモリに転送する（本実施例ではメモリ容量が大きく、すべてのファイルがホストコンピュータの主記憶に格納可能な場合について述べている）次にステップ2において、読み出したファイルに対し特定文字の順序で（例えば漢字コードの垂から）検索していく。そしてステップ3におい

て、もし設定した文字が見つかったなら、ステップ4にて見つかった文字の次の文字を獲得して、さらにステップ5で文書ファイルに対応する識別名称を得る。これらについてステップ6に示すように文書データのすべてについて次々と分析していく。ステップ7では指定した文字を索引として使用可能なようにファイル名（例えば 亜.Psf）を与えて次に引き続く文字と識別名称であるファイル単位の番号を記録していく。そしてもしすべての漢字、片仮名、平仮名について（例え分析が終了したなら連想表は完成する。

【0012】次に図15にて検索の手順を説明する。まずステップ1ではユーザーから入力された検索要求、すなわちj個の検索語ならびに各検索語の論理式が8のホストコンピュータに送られ、ステップ2では各検索語の2文字組の要素に分解（例えば検索語が「松下電器」ならば「松下」、「下電」、「電器」に分解）される。この時メモリ上では索引の対象となりうる文字すべてについての2次記憶上の連想表の物理的な先頭セクタ番地と最終セクタ番地、文書ファイル本文の先頭セクタ番地と最終セクタ番地及びDOSのファイル管理部との対応テーブル、後にでてくる識別名称等の情報を表形式にて持っておく（これらに関する情報は12のファイル管理。そして上記のメモリに先頭文字を参照して連想表の記録セクタに関する物理的な位置情報を得た後、次にステップ3で2文字組の先頭文字をファイル名、すなわち索引とする（例えば松.psf、下.psf、電.psf）連想表の要素を読み込む。次にステップ4にて連想表の要素でそれぞれ第2文字目の文字並び（例えば各々の松.psfでは「下」、下.psfでは「電」、電.psfでは「器」）を検索してその直後に書きこまれている識別名称、すなわちこの場合はファイル名に対応した番号を読み込む。当然のことながら「松下電器」は同一ファイルに存在するのだから各要素は同一の識別名称を持っているはずである。

【0013】さらにステップ5ではj個の検索語に対して検索語間の論理式（例えば「松下電器」&「家電メーカー」ならばすべての検索語すべてが同一の識別名称を持つ）をチェックする。次にステップ6では識別名称に対応するファイルの物理的な記録位置情報をホストコンピュータのファイル管理情報から読み取り、対象となるファイルだけを6のデータメモリに読み込み、5のデータパターン検出回路に検索語をセットして本当にその検索語が存在するファイル（「松下電器」と「家電メーカー」が両方存在するファイル、実際には先に出てきた識別名称で管理）のみをチェックして8のホストコンピュータに送り、もし該当ファイルが存在したらKWIC（キーワードインコンテキスト）法に基づいて検索語を含む行が結果としてユーザに表示される。

【0014】図3は2次記憶に書き込まれている内容であって、11はファイルシステム全体の記録内容の模式図、12はのファイル管理部、15は文書ファイルであ

る本文ファイル群、13はファイルの物理的な位置（物理的な位置は12のファイル管理部から読み取ることができる）から順に番号をふったものを識別名称として与え、実際の物理的な記録位置と大きさ、さらにファイル管理部へのエントリ情報等を表形式にて書き込んだ領域である）また14は連想表であり、実際は各索引別にファイル形式で（例えば亜.pss等）記録されている。なおここでは分かりやすくするため数値で書かれているが、実際にはバイナリ形式で記録する。

【0015】次に図4を用いて別の実施例を示す。図4においては図3の実施例とは2次記憶の内容における連想表の形が異なる。14'の連想表において各索引は20ビット（索引をファイル名で与えた時は3バイトで表され、このうち最初の2バイトを文字列の最初の文字、次の4ビットをアスキー文字1バイトの0からFで表記）で表し、後半の4ビットで第2文字目を大分類している。検索語が入力された時点で文字並びを調べて第1文字目と第2文字目の文字コードの一部を取り出して大分類を行う。連想表の各内容も第2文字目の残りの12ビットをコード化して記述し（3/4と表している）次の第3文字目の情報と、さらに識別名称の組を作って構成している。平仮名や英文字等ではもともとの文字の持つ情報量が小さいため、このようにエントリを1文字以上にしたほうが良い。

【0016】次に図5を用いて別の実施例を示す。図5においては図3及び図4の実施例とは2次記憶の内容における連想表の形が異なる。14''の連想表において索引は1文字であり、表の中身は次の文字1文字に加えさらに次の並びの文字を4ビットで文字を大分類したものとで表している。もし全角文字を12ビットで符号化すれば第2番目の文字と第3番目の文字の4ビット（全12ビットの1/3）を大分類符号として計16ビットで表すことができる。

【0017】次に図6を用いて別の実施例を示す。図6においては図4及び図5の実施例とは2次記憶の内容における連想表の形が異なるが両方の形式を合わせ持ったものである。

【0018】次に図7を用いて別の実施例を示す。図7においてはプリサーチを用いた検索手順を示している。まずステップ1からステップ5までは図15で示した手順と全く同一に処理を進める。ところが最新に登録された文献については、連想表が未作成であると考えられる。これは例えハードウェアを使用したとしても、ある程度の時間がかかるのはやむを得ず、次の登録まで実質的な検索対象とすることができないという問題があった。ステップ6では検索語自体は図15と全く同一に検索回路に複数の検索単語をセットして論理式がORであろうとなかろうと並列OR検索を行い、論理式にAND条件を含む部分についてはOR検索語したのちに得た識別名称が条件どおり存在しているかをチェックして同一

の識別処理を進めるが、検索対象範囲は検索時点で連想表が出来ている範囲とし、その時点で絞り込んだ識別名称をメモリに蓄えておくと共に、ステップ7では検索語は引き続いてまだ連想表が出来ていない文書ファイル本文全てについて検索対象として、検索回路の設定をそのままにして引き続き検索処理を続行する。そしてステップ8で同様にホストは得られた結果を表示する。連想表は例えば週末帰宅前にパッチプログラムをセットしておき、週単位で増加した新規登録文書のデータを連想表に追加していくようにする。次に図8を用いて別の実施例を説明する。図において16は2次記憶のセクタのつながりを模式的に表したものであり、連想表の内容のうち一部をあらわしているところの17は「亜」という文字を索引情報とする、また19は「阿」という文字を索引情報とする表の周辺について物理的なセクターのイメージを記述している。また18は無記録セクタである。このように連想表の内容は各索引文字に対応するデータ毎に連続的な領域に記録され、さらにこれらの間にはギャップを設けている。もし文書ファイルが新規登録され、連想表の内容を追記する場合はホストコンピュータの主記憶上に展開された情報から最終セクタの位置を知ってこれを読み取り変更して書き換えるだけで良く、またもし最終セクタが満杯になったならファイル情報を変更した後、無記録セクタの部分に書き込んでいく。当然この無記録セクタの大きさは一定とすることもできるし、文字の出現頻度に応じて変化させることもできる。それでも追記すべき場所が満杯になってしまった場合には、リンクを張った旨を14の領域に書き込んでおき、とりあえず別の領域にセクタを確保して書き込んでおく。またさらに複数の索引文字の表の情報をあつめてその間に大きなギャップを設けておきここをリンクの飛び先としてもよい。次の実施例としては上記のリンクを張って不連続となったセクタ記録内容をパッチプログラムで修正してまた連続としておくようにもできる。次に図9を用いて別の実施例を説明する。図においては5'の検索回路及び7'の制御回路、6'のメモリ回路の別の形を示している。ここでは本文の全文検索に用いたハードウェアを連想表の文字成分分析に使うように制御回路、及び検索回路の形態を変えている。次の実施例としては図9において特にバイナリ表形式の連想表の検索を行うため、固定の検索パターン長と比較対象としない領域のデータは固定長ジャンプして比較処理をしないようにデータインターバルレジスタというレジスタにデータをセットして6'のメモリに対して検索アドレスを発生する。次に図10を用いて別の実施例を説明する。図においては2次記憶に書き込まれている内容の別の例であって、14''''は連想表であり、実際は各索引別にファイル形式で(例えば亜.pss等)記録されている。ここに「*00」「*FF」と記してあるデータ内容は表の内容のうちとくに識別名称を短いビット数(ここでは8ビッ

ト)で表現したいがために設けたものであって、「*」は未定義文字を表している。そして識別情報を16ビットで表して全部で文書ファイルの数を65536までとした時の上位ビットの桁上がりがあった時にこの未定義文字「*」で表すようにしている。当然のこのながらこの例では上位バイトの「01H」から「FEH」まで該当データが存在せず「00H」と「FFH」にしか文字並びが存在しなかったことを意味しているが、例えば桁上がりがあった時はすべて記述する形式すなわち*01*02*03...*FEHという形にもすることが可能である。また24ビット形式で文書ファイル数を表してこの時の最上位バイトの桁上がりのための未定義文字を「#」として「*」と共に用いることもできる。

【0019】次に図11を用いて別の実施例を示す。図では連想表を作る手順を示している。ステップ1では検索回路を用いる際に文書データをデータメモリに全角文字になおして(例えば英文ならアスキーの大文字になおして)転送する。この後ステップ2では索引となる16ビット文字コードの検出を行う際に、全角文字として有効な範囲の上位8ビットだけを設定ししかも検索は文字単位に行うように設定する。この時およそ下位ビットに含まれる約256文字が検索対象となる。ステップ3及びステップ4では検索ハードを起動して文字検出を行っている。次にステップ5では検索ハードで検出された上位8ビットに加えて、下位8ビットのデータを制御回路内のマイクロコントローラで読み取り、索引としての1文字の確定を行う。またステップ6では文字並びとしての次文字の獲得を同様にマイクロコントローラで行う。次にステップ7では対応する識別名称すなわち文書ファイルに付与された番号を獲得する。ステップ8ではもしすべての文書データについてこの処理が終了したかどうか判定し、ステップ9では対象となる256文字の連想表の内容をマイクロコントローラとホストコンピュータが2次記憶装置を制御して書き込む。ステップ10ではすべての対象となる全角文字が終了したかどうかのチェックを行う。

【0020】次に図12を用いて別の実施例を説明する。この図は検索手順を示している。まずステップ1ではユーザがホストコンピュータを使用してカナ漢字変換プログラムを用いてj個の検索語を入力する。ステップ2ではホストコンピュータから4の処理回路に対し、i番目の検索語を送る。この時4の処理回路では検索語2文字組文字要素の先頭文字を検索の索引としてk1からki-1までの想表検索を行う。これらのi番目の検索語に関する処理データは4の処理回路(検索処理回路)を介して2次記憶に発行されるが、4の処理回路はコマンドを受付た直後に完了信号をホストコンピュータに返し、連想表の検索処理の結果がでるまで待たない。ステップ2'では一つ前のi-1番目の検索語の連想表の検索が終了しているかをチェックしそれも終了していたら

ホストコンピュータは次の検索語のカナ漢字変換を開始する。i-1番目の検索語に関しては存在する可能性の高い文書ファイルを候補文書として4の処理回路から受け取り、該当候補文書としてホストコンピュータの画面に表示する。次にステップ3、ステップ4、ステップ5では4の処理回路内部においてi番目の検索語に関する連想表の検索を行う。さらにステップ5'で全県錯誤の入力が完了しているかのチェックを行ったのち、ステップ6では候補文書の本文検索を行った後、ステップ7で結果を表示する。

【0021】次に図13にて別の実施例を説明する。図13では4の処理回路が1の2次記憶装置の中に組み込まれた形となっている。検索処理は基本的にすべてこの中の処理回路とホストコンピュータで行われ、効率的に処理を行うことができる。

【0022】次に図14にて別の実施例を示す。図14は図2と比較するとステップ5'とステップ5''の部分異なる。ステップ5'では調べた次文字と識別名称の組のうち重複しているものをチェックする。そしてステップ5''では重複するものを削除して記録する。このことにより、連想表の容量を小さくすることができる。

【0023】別な実施例としては当然上記のこの削除した組の数をチェックしておき、各索引文字単位に次の文字並び単位に何個削除したかのデータを取っておくことにより2文字組の出現データを得ることができ、これをソーティング後に2次記憶装置の一部に書いておくことができる。

【0024】また次に別の実施例では、連想表における識別名称を例えば64KB以下とし、64KBとなるように複数の文献をまとめ、また64KBを越える分については、文の区切りを調べた上で64KB以下となるように分割しこれらの単位に識別名称を与える。この情報は2次記憶の12の部分に書き込んだのち、主記憶に取り込み、検索語間の論理式については本文検索時に元の各文献単位毎にチェックを行う。

【0025】また別の実施例においては、検索語として「塩化ビニール」という検索語を選んだ時、「塩化」「化ビ」「ビニ」「ニール」の組み合わせ5つ全部の論理積を取らずに例えば平仮名については3つまで存在したら良いという条件でチェックし、また例えば片仮名については、表記の揺らぎがあるためより条件を緩くする、あるいは揺らぎを含めて本文検索を行う等の処理ができる。またプリサーチ時に条件を緩くした状態での存在したところの候補検索語となりうる文字列の組み合わせを一旦ユーザに表示して選択してもらい、最終的な本文検索を行うようにすることもできる。

【0026】なお本実施例では、1台の2次記憶装置を用いて説明したが、特にこれらは1台に限定されることなく、検索回路、処理回路も1台から複数台で用いることができる。また最終的な表示もKWIC方式しか示

していないが、他の方式やあるいはまたユーザの指示によりヒット行にカーソルを持ってきたら文書ファイル生データを表示する等の処理も可能である。また英語等の外国語を対象とした場合についても同様にできることはいうまでもない。

【0027】また候補となった文献の件数を画面に表示して、条件を満足した文献数が適当な数になるようにユーザに次の検索語と論理式を選んでもらうこともできる。

【0028】また連想表の各索引に対応する内容を取り込んだり転送している間に、次のトラックにシークする等の高速化処理（プリフェッチ）ができるのも当然である。

【0029】なお本発明は上記実施例に限定されるものではなく、本発明の主旨に基づいて種々の変形が可能であり、これらを本発明の範囲から排除するものではない。

【0030】

【発明の効果】本発明によれば、次のような効果を得ることができる。

【0031】検索のプリサーチとして作成する連想表は特許文書を対象とし、2文字組で1文字を索引文字として使用した場合で本文の約1/5、新聞で1/3程度であった。このようにする比較的小さなプリサーチファイルを用いながらもプリサーチファイル全部を検索対象としないため非常に高速にしかも効率良く全文検索できる。

【図面の簡単な説明】

【図1】本発明の一実施例における検索装置の構成図

【図2】同検索装置における手順を示すフローチャート

【図3】同検索装置におけるファイルの構成図

【図4】同検索装置におけるファイルの構成図

【図5】同検索装置におけるファイルの構成図

【図6】同検索装置におけるファイルの構成図

【図7】同検索装置における検索手順を示すフローチャート

【図8】同検索装置における2次記憶装置内連想表の記録状態の説明図

【図9】本発明の一実施例における検索処理回路の構成図

【図10】同検索装置におけるファイルの構成図

【図11】同検索装置における連想表作成手順を示すフローチャート

【図12】同検索装置における検索手順を示すフローチャート

【図13】本発明の一実施例における検索装置の構成図

【図14】同検索装置における連想表作成手順を示すフローチャート

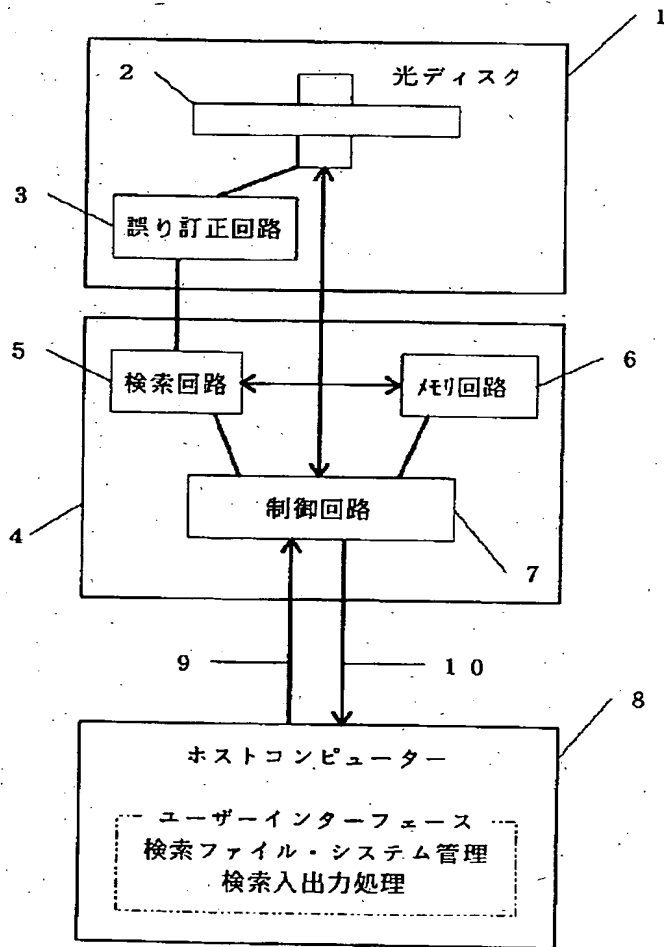
【図15】同検索装置における検索手順を示すフローチャート

【符号の説明】

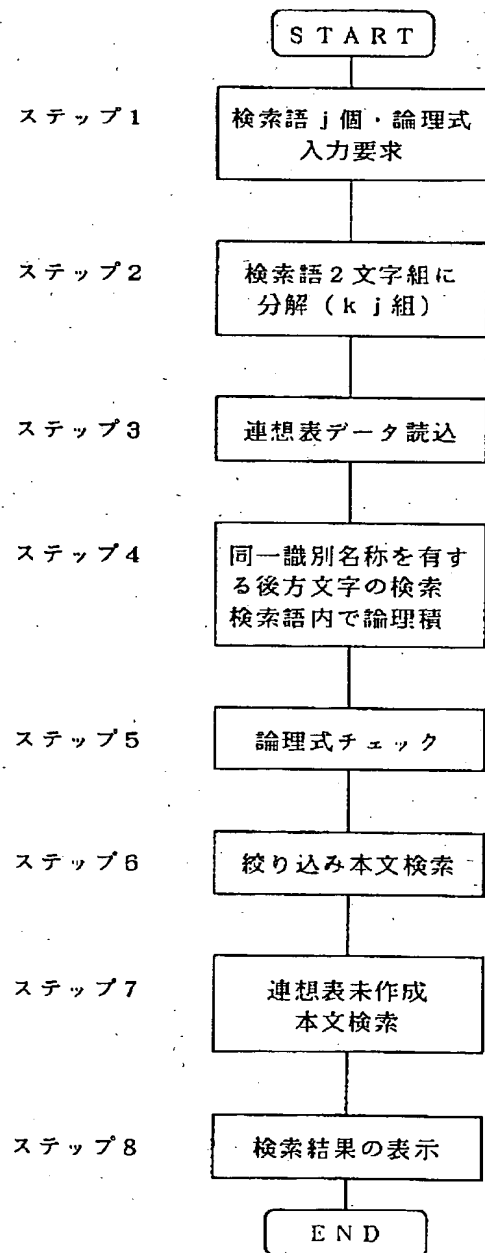
- 1 2次記録装置
- 2 記録媒体（例えば光ディスク）
- 3 誤り訂正回路
- 4 処理回路
- 5 文字列検索回路
- 6 データメモリ回路
- 7 制御回路

- 8 ホストコンピューター
- 9 検索要求
- 10 検索結果
- 11 ファイルシステム全体の記録内容の模式図
- 12 ファイル管理領域
- 13 情報記録領域
- 14 連想表領域
- 15 本文ファイル領域

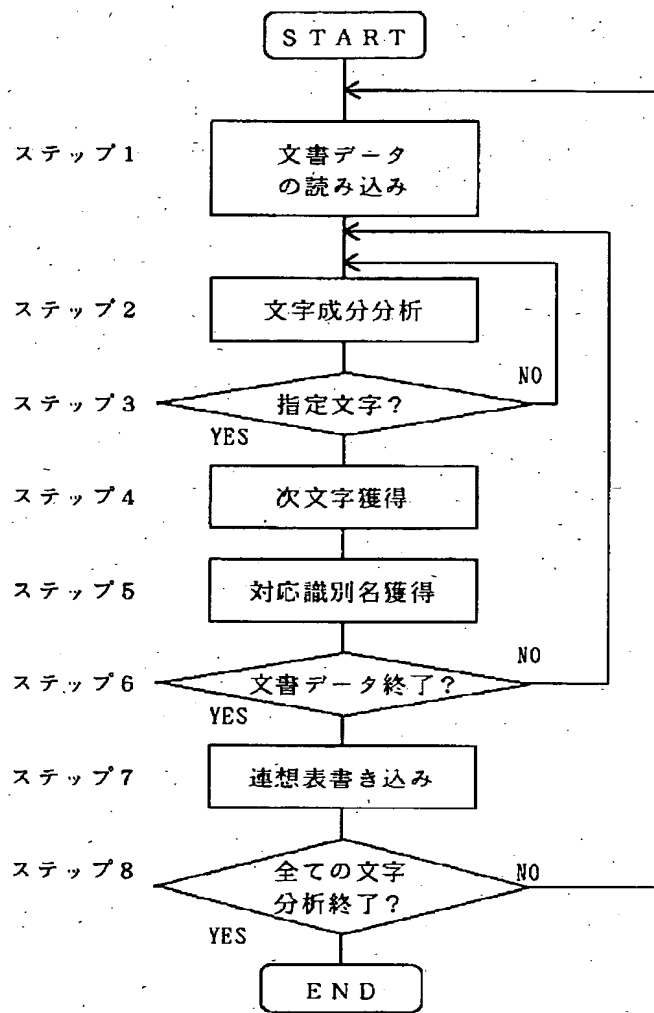
【図1】



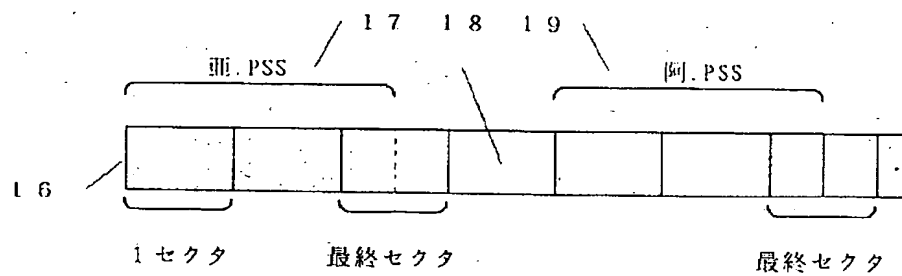
【図7】



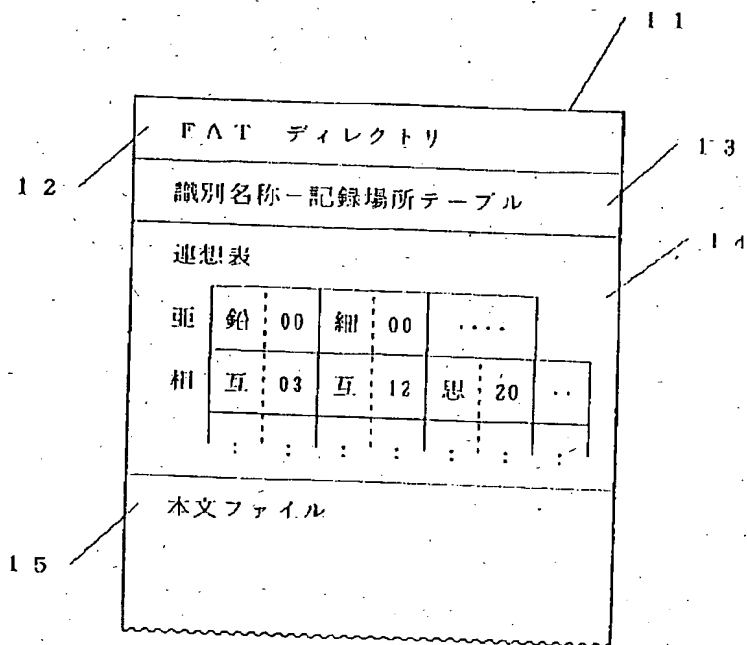
【図2】



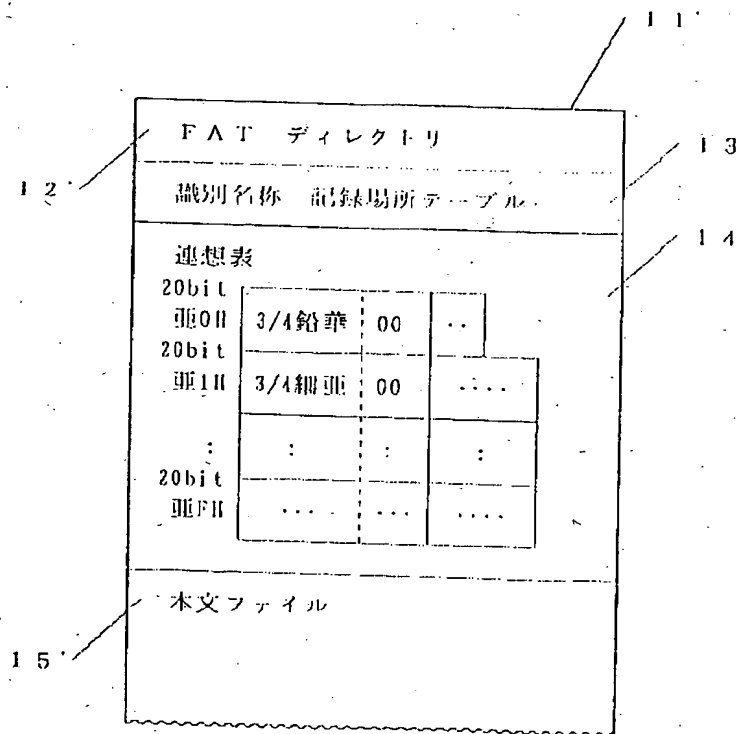
【図8】



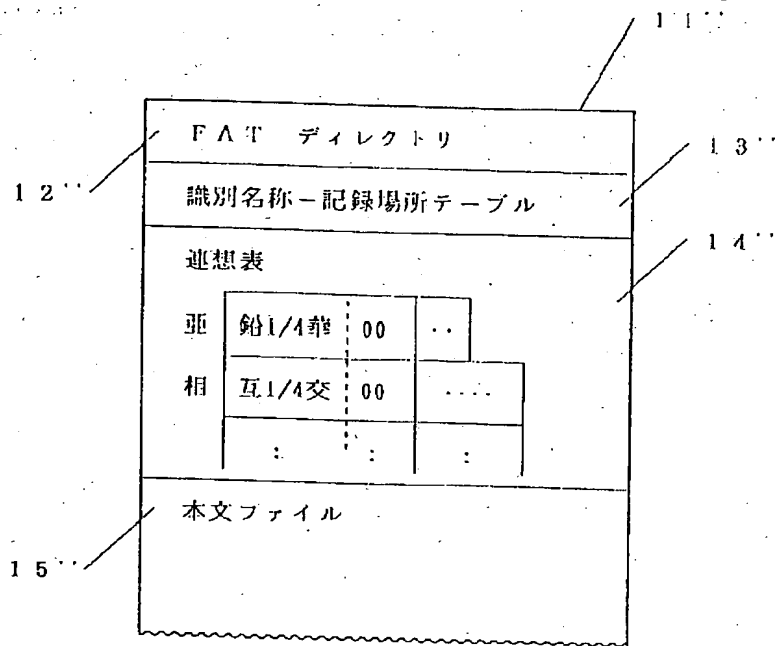
【図3】



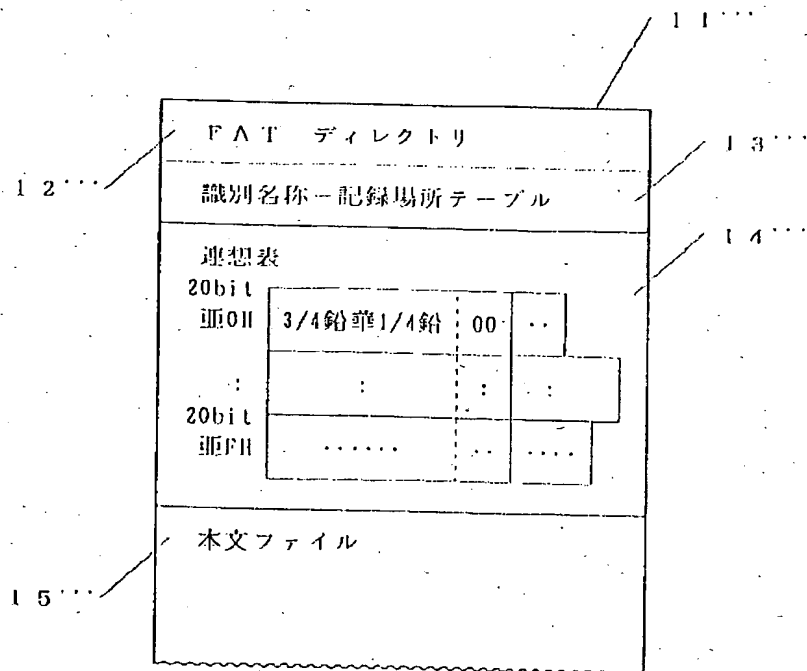
【図4】



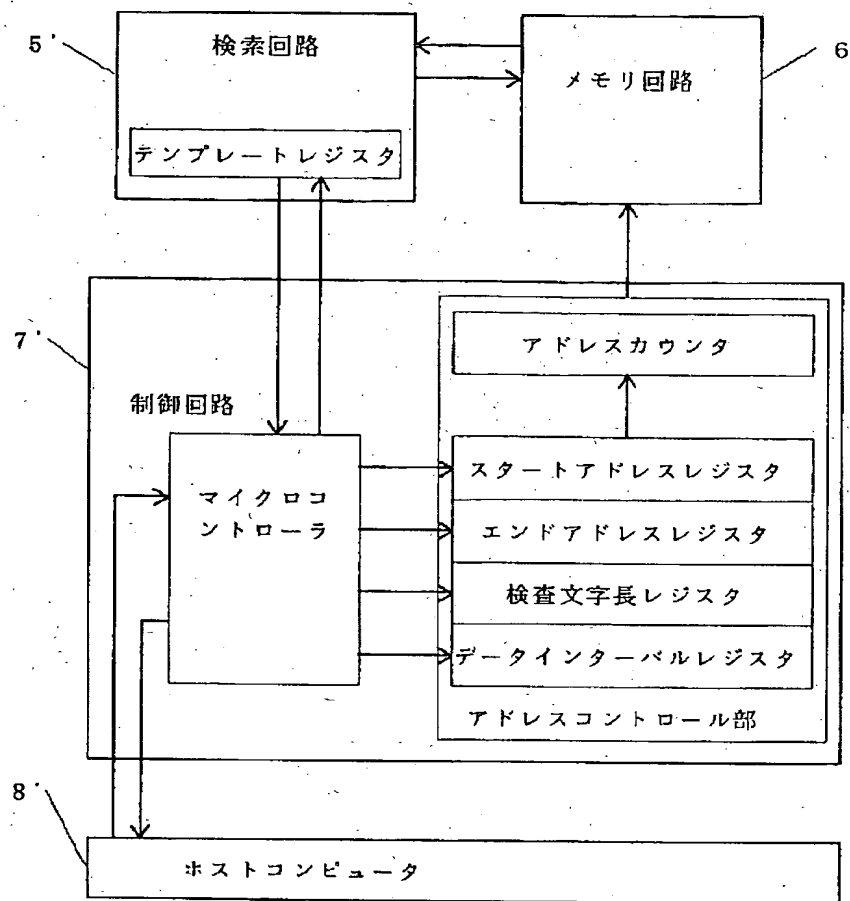
【図5】



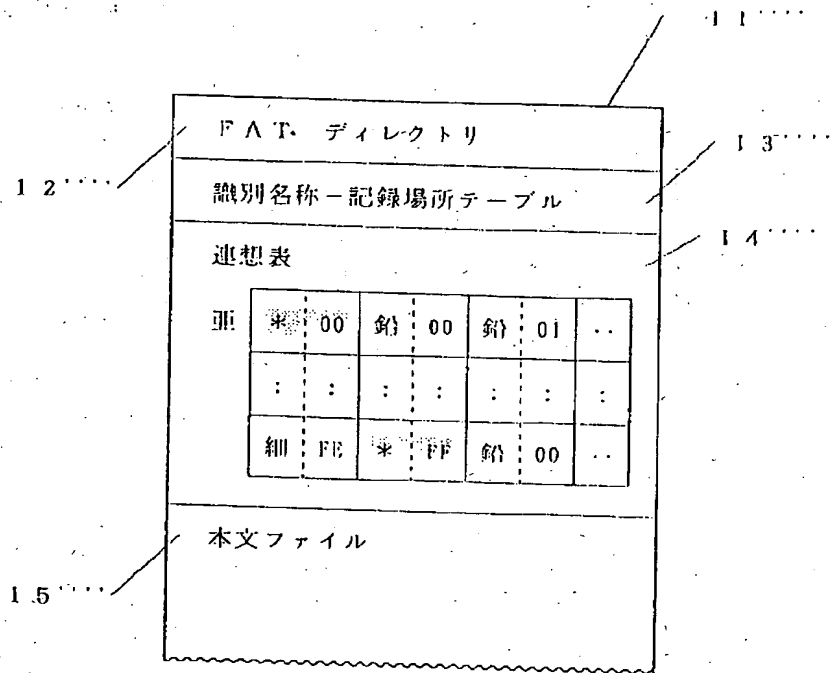
【図6】



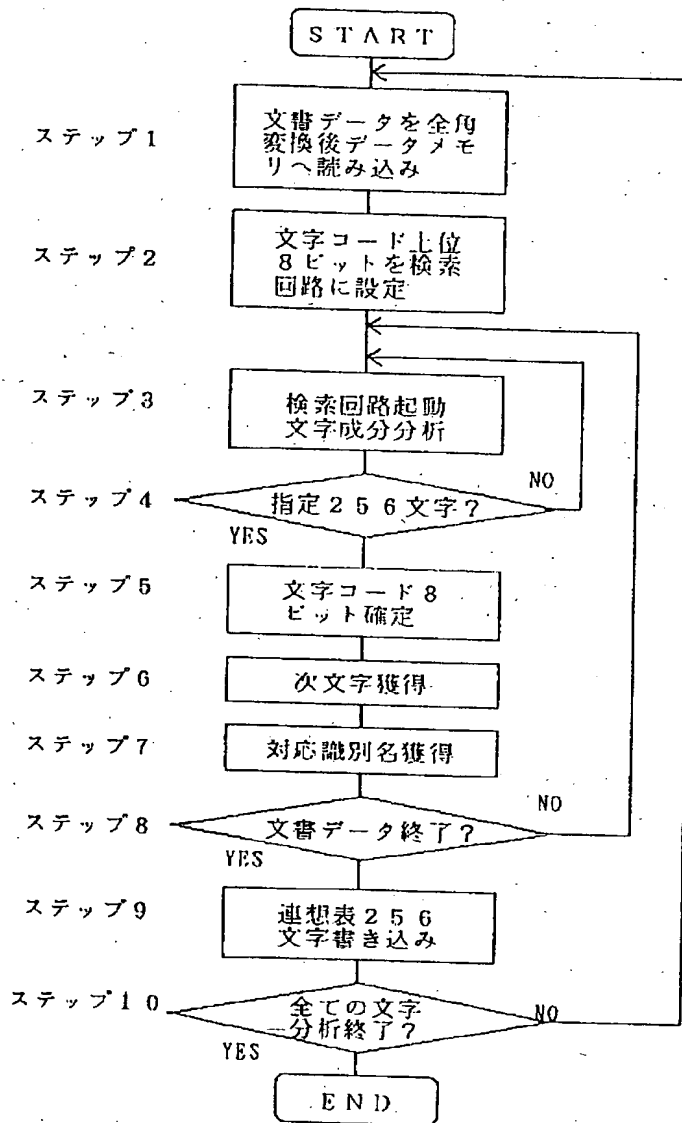
【図9】



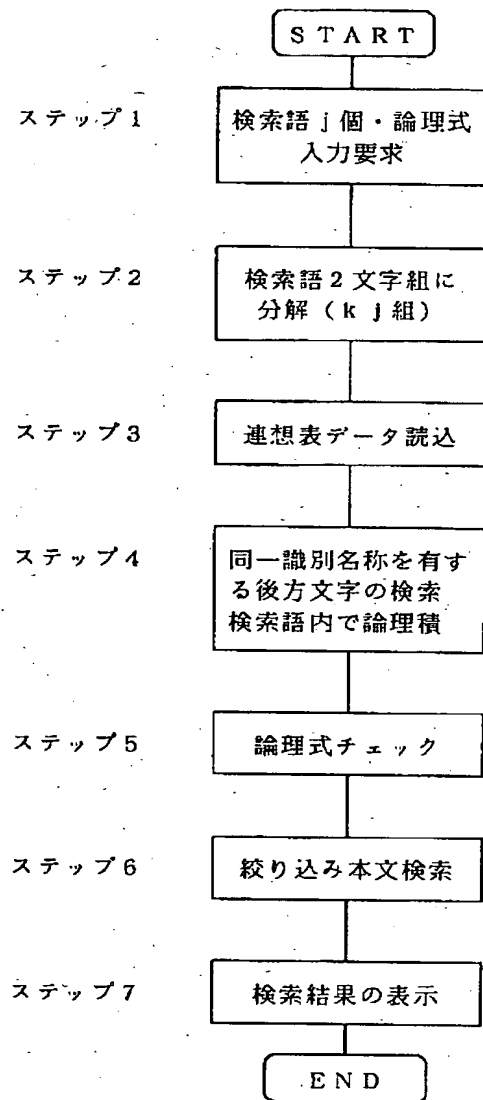
【図1-0】



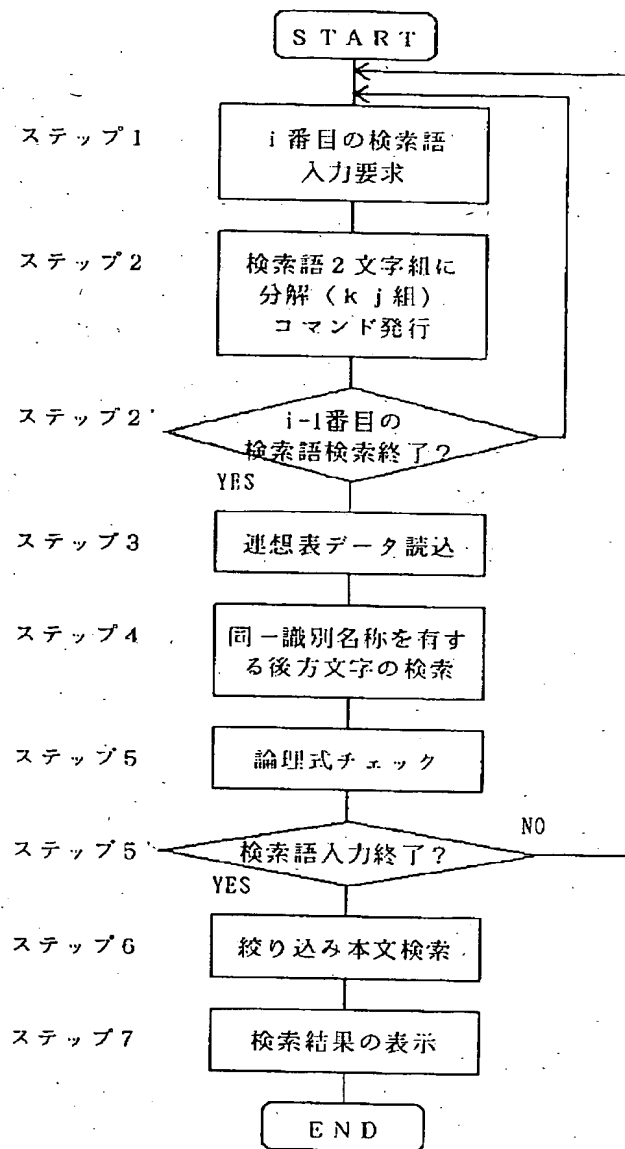
【図11】



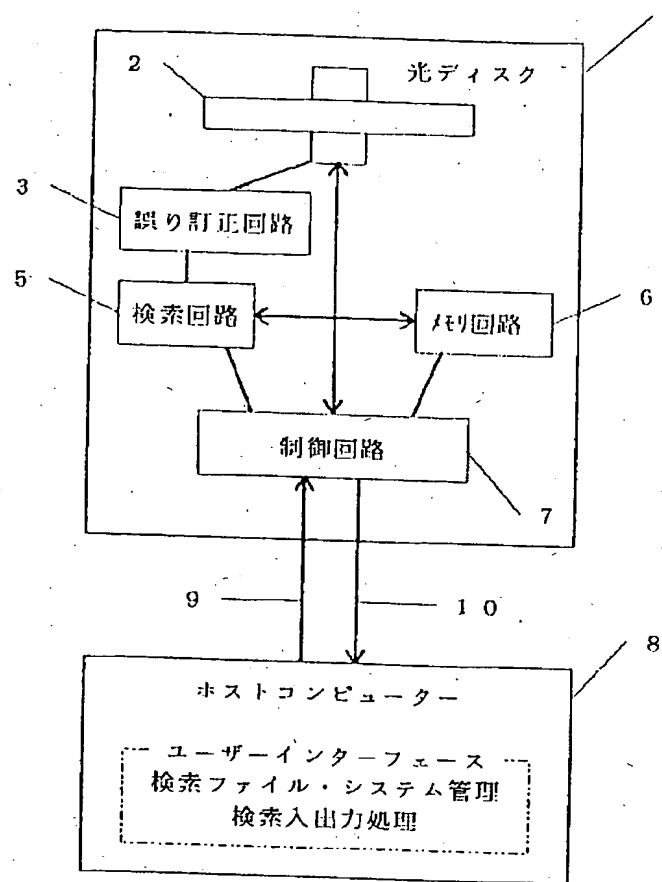
【図15】



【図12】



【図13】



【図14】

